Preference vs. Authority: A Comparison of Student Searching in a Subject-Specific Indexing and Abstracting Database and a Customized Discovery Layer

Sarah P.C. Dahlen and Kathlene Hanson

Discovery layers provide a simplified interface for searching library resources. Libraries with limited finances make decisions about retaining indexing and abstracting databases when similar information is available in discovery layers. These decisions should be informed by student success at finding quality information as well as satisfaction with search interfaces. Students executed searches in two discovery layer configurations and an indexing and abstracting database. While students reported a preference for discovery layers, the articles selected from the indexing and abstracting database were more authoritative. These results illuminate the relative strengths of these tools, informing decisions about resource allocation and discovery configuration.

Introduction

Libraries are at a crossroads when it comes to providing access to bibliographic information and full text of library resources. The expectation of students and faculty for a "simplified, fast, all inclusive, and principally online research experience"[1] has been addressed in part by the introduction of discovery layers, which have been widely adopted by libraries.[2] Librarians have also stressed "the need for a single point of entry,"[3] which has likely contributed to the more-than-doubled use of discovery layers in libraries observed from 2010 to 2011,[4] despite the significant cost, which can be prohibitive for smaller institutions. Students have indicated their satisfaction with the usability of discovery layers;[5] and, while many prefer the simplified interface, these tools return a large number of search results that can include popular as well as scholarly sources. Some librarians are skeptical about searches yielding too many and irrelevant results,[6] thus putting a greater burden on the student searcher to sort through and evaluate information. If students are disinclined to go beyond the first page of search results or if they lack the skills to properly evaluate information

Sarah P.C. Dahlen is Research and Instruction Librarian and Kathlene Hanson is Electronic Resources Coordinator in the Tanimura and Antle Family Memorial Library at California State University, Monterey Bay; e-mail: sdahlen@csumb.edu, khanson@csumb.edu. ©2017 Sarah P.C. Dahlen and Kathlene Hanson, *Attribution-NonCommercial (http://creativecommons.org/licenses/by-nc/4.0/) CC BY-NC.*

878

sources,[7] they may select sources that are lower quality with regard to their authority, currency, or relevance.

In indexing and abstracting bibliographic databases, some of this sorting and evaluation has already occurred. Many academic libraries provide subject-specific indexing and abstracting databases, and academics continue to promote them to students, assuming that the depth of coverage and the traditional indexing of citations mean more relevant search results. As discovery layers rise in popularity among users, however, librarians may be tempted to cut indexing and abstracting databases in the face of budget constraints. To determine whether subject-specific indexing and abstracting databases continue to provide added value for libraries that are now investing in discovery layers, this investigation was designed to compare how each type of tool affects students searching for information.

The goal of this study is to shed light on the following questions:

1. Does the type of library search tool used by students affect the quality of information sources that students select?
2. Does the chosen tool affect student satisfaction with the search experience?

Literature Review

Federated searching was the predecessor to web-scale discovery, offering students a way to search across multiple databases simultaneously, and was once "touted as the library world's answer to Google."[8] However, as noted by Rose-Wiles and Hofmann, federated searches were "difficult to customize and cumbersome to use."[9] In addition, these were never really all-inclusive, as federated searching interfaces primarily searched content from databases to which institutions subscribed. Discovery layers have largely supplanted federated searching, creating direct access to indexed content without the need to search against other databases. Web-scale discovery products index a vast number of resources in many formats, allowing users to search across collections with a single search box.[10]

A number of studies have indicated high levels of student satisfaction with web-scale discovery layers. For undergraduate users who have grown up with Google, a web-scale single search box has become the expectation, and some librarians believe that discovery layers will bring these users back to the library.[11] Lundrigan, Manuel, and Yan found that the majority of respondents to their user satisfaction survey were "very" or "moderately" satisfied with the Summon discovery layer, though this was less true for graduate students and students in certain disciplines.[12] Student participants in Asher, Duke, and Wilson's study indicated a preference for discovery tools and Google Scholar over traditional library tools such as the catalog and databases.[13] As Gross and Sheridan conclude, "there is no doubt that the students found [the web-scale discovery tool] an easy way to get results and probably easier than the various options they were previously faced with."[14]

While students seem to prefer the simplified search interface, discovery layers also have disadvantages. Gross and Sheridan observe that "it appeared that the students were not able to fully understand the results they obtained, or where those results were coming from."[15] Rose-Wiles and Hofmann summarize commonly cited disadvantages of the web-scale discovery services, including "lack in precision yielding too many results and irrelevant results, gaps in coverage, problems limiting or refining results, and general information overload," all of which may contribute to students not fully understanding the results list.[16] Further, Howard and Wiebrands found that librarians were not as satisfied with the Summon discovery layer as they were with more traditional databases,[17] even while students expressed their preference for the single search box's ease of use in another study at the same institution.[18] The librarians were

concerned about coverage and "dumbing down" of searching, as well as whether students were finding the most useful resources. While discovery layers offer facets allowing users to limit their results, some studies have indicated that students may have trouble limiting or refining results when using these tools.[19] Asher, Duke, and Wilson state that students showed "a marked inability to effectively evaluate sources and a heavy reliance on default settings."[20] Additionally, in a study of gaze behavior, Kules and Capra report that, for students searching library resources, only about 10%–30% of their gaze time was on facets,[21] reinforcing the assertion that users tend to rely on the default configuration.

Theory and research on information retrieval behavior helps inform the processes through which database users make decisions about source selection. There are numerous established models and theories related to information behavior and the involved activities of encountering, needing, finding, choosing, and using information.[22] Project Information Literacy has found that efficiency and predictability are key factors motivating college students' information-seeking behavior.[23] Agosto investigated web-based decision making through the lens of Simon's behavioral decision-making theories of bounded rationality and satisficing and found that participants demonstrated satisficing behaviors, or accepting convenient options that are merely adequate, in deciding which resources to use.[24]

Asher, Duke, and Wilson conducted an investigation of how discovery layers function for students, comparing two discovery layers (EBSCO Discovery Service and Summon) along with Google Scholar and traditional library catalogs and databases.[25] Student participants at two universities executed searches in one of these tools and selected information sources based on research prompts. These sources were evaluated using a holistic, home-grown rubric that combined elements of authority, relevance, and currency. The researchers found that EBSCO Discovery Service outperformed other tools in the quality of sources chosen by students, while Summon, the other discovery layer included, did not have a statistically significant difference in performance compared to the library catalog and databases. The authors concluded that one of the most important factors in the sources chosen by students is the way that sources are ranked in the results list.

Another approach to search tool evaluation is to assess the usability of different platforms. For example, both Cordes[26] and Gross and Sheridan[27] investigated the usability of various search tools. Cordes had undergraduates perform search tasks as part of a course and then administered a survey gauging participant perceptions of the usability of search tools, including ease of use and usefulness. While the study found a correlation between perceived ease of use and usefulness, search engines received higher rankings for ease of use than databases, but databases were rated as the most useful, indicating that users differentiate between these two variables.[28] Gross and Sheridan investigated the navigation, ease of use, and search result quality of Summon by having students complete research tasks in this discovery layer. The researchers found that having a single search box simplified navigation for students, though they did have trouble interpreting the results list and understanding the different formats included (such as journal article versus newspaper article). The authors concluded that, while students can find information in Summon, it is not clear whether they can evaluate it.[29]

While some studies use search results as a relative measure of how well students complete research tasks in various types of search tools,[30] other studies concentrate on analyzing the qualities of the search results themselves regardless of the search tool used. Hovde suggests that citation analyses of student paper bibliographies "provide a flexible, non-invasive, time-efficient assessment forum for the documentation of student library use."[31] Leeder, Markey, and Yakel surveyed a number of articles that

employ citation analysis, noting that this technique has often been used for library assessment purposes.[32] Their review of the literature found that none of the published bibliographic evaluation criteria covered a sufficiently wide variety of source types and that many tools were qualitative and not tested for validity or reliability. Subsequently, they created their own taxonomy for rating student bibliographies and the individual citations they contain. This taxonomy is format-neutral and was revised through an iterative process, including input from instructional faculty and testing for interrater reliability.[33] Leeder, Markey, and Yakel's taxonomy was employed in our study to measure the authority of information sources.

Methodology

Given the popularity of discovery layers and the scarcity of studies investigating how effectively they are employed by students, we designed our research to explore how discovery layers compare to subject-specific indexing and abstracting databases in student searches. We wanted to see if the student preference for discovery layers over traditional library databases, as described in the literature, was reflected among students at our institution. More important, we hoped to determine whether the type of library search tool used by students affects the quality of information sources that students select.

To gain a better understanding of the effect that different types of search tools have on the information selected by students, we conducted our research on search tools that we hoped would be generally representative of the two categories of interest: discovery layers and subject-specific indexing and abstracting databases. Social Sciences Abstracts was chosen as the indexing and abstracting database for two reasons: 1) it was one of the databases considered for elimination when our library faced financial constraints on electronic resources, making an understanding of its utility to students of particular use for our library; and 2) while not specific to a single discipline, its scope mirrors that of our interdisciplinary Social and Behavioral Sciences major, making it a good match for students doing library research in this area.

Serial Solutions' Summon was the discovery layer employed for this project because it is what our library currently provides. Rather than the standard Summon search interface, our university system created its own search interface, called Xerxes, which works with the Summon API. While we hope that the results of this study are somewhat generalizable to discovery layers as a category, it should be noted that the specific interface of any discovery layer may have an impact on how effectively it is employed by students.

The default configuration of a discovery layer will affect how students experience it, and libraries have some flexibility in determining the default settings. For example, libraries may choose whether to include nonscholarly sources, such as newspapers and magazines, in the initial set of search results. These parameters may be changed by the searcher once s/he is on the results page; but, because many students do not change the parameters,[34] the default configuration is an important element in determining which results are displayed for student users. Our library's default configuration of Summon includes only citations for which we subscribe to the full text, and it includes scholarly as well as popular and trade publications. We will refer to this configuration as "default Summon."

To explore further the effects that these default parameters have on student use of discovery layers, we decided to include a second configuration of Summon, one that would more closely replicate the experience of searching a subject-specific indexing and abstracting database such as Social Sciences Abstracts. For this project, we created a Summon search box whose default configuration was set with limitations on the formats and subjects of the articles returned. The format was set to exclude newspapers and to

include journal articles, magazine articles, book reviews, dissertations, book chapters, trade publication articles, and conference proceedings. The subject area limiter was set to include the social sciences, such as history, archaeology, anthropology, women's studies, political science, psychology, and sociology. This configuration will be referred to as "pre-scoped Summon." It should be noted that, for both of our Summon configurations, these parameters apply to the initial search; once students view the initial results, they have the option of changing the filters (for format, subject, and so on) that have been applied.

Study Population

Because the indexing and abstracting database used in this study was Social Science Abstracts, participants were selected from the population expected to benefit most from this database: upper-division undergraduates majoring in Social and Behavioral Sciences. Participants were chosen solely on the basis of enrollment at the university, major, and year in school. Participants were asked to report their GPA so that their level of academic preparedness could be controlled when analyzing the results.

Sampling Design

The population of interest was approximately 192 students (upper-division majors in Social and Behavioral Sciences in Spring 2015). Our sample of 50 represents 26 percent of the population of students from which it was drawn. To recruit these participants, a call for volunteers went out to the entire population of interest. The first 50 volunteers were accepted as participants, and they were given gift cards as incentives for their participation. Because this was not a random sample, it may be that the study participants are not fully representative of the population. A total of 31 participants were female and 19 were male, which, coincidentally, was the exact proportion of females to males in the student population in the semester this research was conducted. Participant age was not considered a relevant variable and not recorded.

Data Collection

Student participants were asked to execute searches for articles on a given topic in each of the three library search interfaces described above: Social Sciences Abstracts, default Summon, and prescoped Summon. The searching order of the three tools was randomized to control for order effects, and the searches were in response to specific prompts emulating a research assignment of the type that we see at our university, where the prompt is broad enough to allow students some flexibility (see appendix A). The topics of the prompts were similar to topics chosen by former students in the Social and Behavioral Sciences major for their senior capstone projects. We asked participants to choose the two "best quality" articles for the given topic in each tool, with the criteria for "best quality" being left to the student's discretion. The researcher asked questions during the searches about the participant's experience, including why the participant chose specific articles, and asked additional questions after the searching was completed to gauge participant reaction to the search tools and delve further into motivation for source selection. A screencast of participants' searches and their answers to these questions was captured. An evaluative survey was administered at the conclusion of the searches to quantify student experience and preference for the tools (see appendix B). Students were given as much time as they needed to complete the searches, and the length of the sessions ranged from 26 to 84 minutes, with an average of 46 minutes.

The articles selected by students from each search tool were later evaluated by two librarians not present during the searches to eliminate the potential bias of knowing the tool with which articles were found. These articles were evaluated based on three primary criteria: authority, relevance, and currency.

Authority

The first criterion was scored using the faceted taxonomy developed by Leeder, Markey, and Yakel[35] for evaluating student bibliographies. The authors describe their taxonomy as measuring article "quality" through five different facets: 1) information format (such as blog, encyclopedia, scholarly journal); 2) literary content (such as editorial, working paper, research article); 3) author identity (examples: unknown, corporate, academic); 4) editorial process (for instance: self-published, editorial staff, peer-reviewed); and 5) publication purpose (such as commercial, government, higher education). We view these facets as primarily addressing the authority of the publication and prefer to use that term rather than the more general "quality" descriptor employed by the authors.

Relevance

We thought it was important to include some measure of the relevance of the articles selected, as a highly authoritative article is not valuable for a student's purposes if its topic is irrelevant. Finding an existing measure for the relevance of the articles selected by our participants proved to be more difficult than anticipated. Leeder, Markey, and Yakel[36] reviewed the literature for criteria used for bibliography evaluation; and, of the 29 sets of published criteria they found, only four addressed relevance. Of these four, none were appropriate for our purposes, as they either focused on the mix of citations in an entire bibliography (rather than individual citations) or they failed to define the standards by which relevance was to be measured. Part of the dearth of relevance measures for bibliographies or citations may be a result of the difficulty of determining relevance from these artifacts alone. In the context of an essay or research paper, a student is expected to demonstrate through the narrative how an article cited is related to the topic at hand. Evaluating a citation without this narrative context makes it challenging to determine the student's intentions for connecting the article to the topic. Nonetheless, we thought it important not to ignore the issue of relevance and created our own rubric to measure it (see appendix C). This rubric was developed by the researchers and vetted for validity by other librarians at our university.

Currency

The publication date of the articles was our final criterion of interest, as we wished to determine whether any of the search tools would bias students toward selecting more or less current articles. The year of publication was recorded for each article selected by students.

Results

This study set out to address the effects of several library search tools on student satisfaction to see whether trends reported elsewhere[37] were true at our university. It also aimed to go beyond student preference to investigate whether the type of library search tool used by students affects the quality of information sources that students select, building on the work of Leeder and Asher.[38] Our results shed some light on both of these questions.

Indicators of Interrater Reliability

The simplest way to estimate interrater reliability is to look at the percent of exact agreement among raters. Using this method yielded the results presented in table 1. As shown, for articles found using default Summon, the percent of exact agreement between the two raters ranged from a low of 80 percent (for the relevance rubric) to a high of 100 percent (for taxonomy facet 3). Among those found as a result of using prescoped Summon, the percent of exact agreement ranged from a low of 78 percent

(relevance rubric) to a high of 100 percent (taxonomy facet 3). Among those articles found as a result of using Social Sciences Abstracts, the percent of exact agreement ranged from 70 percent (relevance rubric) to 100 percent (all taxonomy measures). Thus, the least amount of agreement was generated using the relevance rubric and the highest degree of agreement for the taxonomy's facet 3 (author identity).

TABLE 1 Percent Exact Agreement among Raters by Article Number and Search Tool						
	Default Summon		Prescoped Summon		SSA	
	Article 1	Article 2	Article 1	Article 2	Article 1	Article 2
Facet 1: Information Format	90%	92%	96%	94%	100%	100%
Facet 2: Literary Content	94%	96%	92%	98%	100%	100%
Facet 3: Author Identity	100%	98%	100%	100%	100%	100%
Facet 4: Editorial Process	90%	98%	96%	96%	100%	100%
Facet 5: Publication Purpose	94%	94%	88%	92%	100%	100%
Relevance Rubric	80%	80%	84%	78%	72%	70%

Student Satisfaction with Search Tools

We generated descriptive statistics to analyze the results pertaining to student preference for each of the search tools, a topic addressed on the student survey. Table 2 presents the distribution of responses to the item asking students to rate how easy each tool was to use on a scale from 1 to 5, with 5 indicating the greatest level of ease. Students rated the default Summon as being easiest to use, with 54 percent (n = 27) of students endorsing a rating of "5," compared to 38 percent (n = 19) and 40 percent (n = 20) of students endorsing the same rating for prescoped Summon and for Social Sciences Abstracts.

TABLE 2 Ease of Learning Tool: Student Rating by Search Tool						
	Default Summon		Prescoped Summon		SSA	
Learning to use this tool was easy. Do you agree with this statement?	Count	%	Count	%	Count	%
1 (strongly disagree)	2	4%	2	4%	2	4%
2	0	0%	2	4%	6	12%
3	4	8%	4	8%	5	10%
4	17	34%	23	46%	17	34%
5 (strongly agree)	27	54%	19	38%	20	40%
Total	50	100%	50	100%	50	100%

Table 3 depicts the level of satisfaction that students reported with the results produced by each of the three search tools, with larger values representing higher satisfaction. Students were most satisfied with default Summon and least satisfied with Social Sciences Abstracts.

TABLE 3						
Satisfaction with Results: Student Rating by Search Tool						
	Default Summon		Prescoped Summon		SSA	
How satisfied were you with the search results?	Count	%	Count	%	Count	%
1 (not satisfied)	2	4%	2	4%	1	2%
2	1	2%	4	8%	7	14%
3	6	12%	3	6%	8	16%
4	13	26%	18	36%	14	28%
5 (very satisfied)	25	56%	23	46%	20	40%
Total	50	100%	50	100%	50	100%

Looking at which search tool students felt would be most useful in their coursework, the results presented in table 4 are consistent with the previous two questions. That is, students rated default Summon as being potentially more useful in their courses compared to both prescoped Summon and Social Sciences Abstracts.

TABLE 4						
Usefulness for Course Work: Student Rating by Search Tool						
	Default Summon		Prescoped Summon		SSA	
I would find this tool useful in my coursework. Do you agree with this statement?	Count	%	Count	%	Count	%
1 (strongly disagree)	2	4%	3	6%	2	4%
2	1	2%	2	4%	3	6%
3	5	10%	4	8%	10	20%
4	14	28%	17	34%	11	22%
5 (strongly agree)	28	56%	24	48%	24	48%
Total	50	100%	50	100%	50	100%

When we look at the data in tables 2, 3, and 4 in a different way, by counting the number of students responding with either a four or a five on the five-point scale, we can see that the numbers for default Summon and prescoped Summon cluster together (44 and 42 for ease of learning, 38 and 41 for satisfaction with results, and 42 and 41 for usefulness for coursework). There is a gap between these numbers and the number of students giving a 4 or 5 to Social Sciences Abstracts (37 for ease of learning, 34 for satisfaction with results, and 35 for usefulness for coursework). This reiterates the student preference for the Summon configurations over Social Sciences Abstracts.

Finally, two items on the survey spoke directly to student preference by inquiring, "which tool would you most/least prefer to use?" Table 5 illustrates that the number of students rating each tool as most useful is almost equal. However, it also indicates that the percentage of students who least prefer prescoped Summon and Social Sciences Abstracts is double the percentage who least prefer default Summon, suggesting a preference for default Summon that is consistent with the data presented in the three previous tables.

	Most Preferred		Least Preferred	
	Count	%	Count	%
Default Summon	17	34%	10	20%
Prescoped Summon	16	32%	20	40%
SSA	17	34%	20	40%
Total	50	100%	50	100%

TABLE 5
Most and Least Preferred Tool

Quality of Articles Selected by Student Participants

Recall that, for each of the three search tools, study participants were asked to choose two articles, yielding a total of six articles selected by each student. Our raters scored each of these six articles on several measures (the five facets of Leeder, Markey, and Yakel's taxonomy,[39] resulting in an average taxonomy score, and the relevance score). The two raters scoring each of these articles were unaware of the search tools from which they were chosen.

Tables 6 and 7 show the mean rating scores observed for each of these variables. Table 6 includes all the taxonomy ratings, the possible values of which ranged from a low of 1 to a high of 4. Table 7 records scores from the relevance rubric, whose possible value ranged from a low of 0 to a high of 3. The mean values presented in tables 6 and 7 indicate that the articles were scored very highly by both raters in all measures, where no values less than 3.50 for the taxonomy and 2.44 for the rubric are observed.

The research question examined in this section is whether the type of library search tool used by students affects the quality (authority, relevance, and currency) of information sources that students select. To address this question, each pair of tools was compared.

We followed procedures to condense the number of variables (see appendix D). To control for an inflated chance of obtaining statistically significant results due to the large number of comparisons being run, we used an adjusted, more rigorous P value of .007 (.05/7 comparisons) to establish statistical significance.

The average scores obtained for each of the five taxonomy facets (all measuring aspects of authority) and an average taxonomy score are presented in figure 1. Stars represent statistically significant differences meeting the $P < .007$ criteria. Downward arrows represent differences that meet the $P < .05$ uncorrected criteria, which are less conservative. The color of the symbols represents the tool for which the comparison is being made. For example, looking at the rating scores associated with facet 5, a grey arrow above the value of 15.9 associated with Social Sciences Abstracts indicates that this value is statistically larger than the 15.4 value assigned to articles obtained using default Summon at the $P < .05$ level.

	Default Summon				Prescoped Summon				Social Sciences Abstracts			
	Article 1		Article 2		Article 1		Article 2		Article 1		Article 2	
	Rater 1	Rater 2	Rater 1	Rater 2	Rater 1	Rater 2	Rater 1	Rater 2	Rater 1	Rater 2	Rater 1	Rater 2
Facet 1: Information Format	3.82	3.90	3.78	3.86	3.96	3.92	3.90	3.96	4.00	4.00	3.96	3.96
Facet 2: Literary Content	3.71	3.71	3.82	3.78	3.90	3.86	3.90	3.88	3.98	3.98	3.96	3.96
Facet 3: Author Identity	3.97	3.97	4.00	3.96	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00
Facet 4: Editorial Process	3.93	3.92	4.00	3.94	3.96	4.00	4.00	3.96	4.00	4.00	4.00	4.00
Facet 5: Publication Purpose	3.85	3.90	3.82	3.84	3.84	3.96	3.88	3.96	4.00	4.00	3.96	3.96
Average Taxonomy Score	3.86	3.88	3.90	3.90	3.98	3.96	3.98	3.98	4.00	4.00	3.98	3.98

TABLE 6
Average Taxonomy Rating Scores by Article and Rater (score range 1–4)

Looking just at differences that are statistically significant at the more rigorous $P < .007$ standard, we see that articles obtained using Social Sciences Abstracts received higher scores on taxonomy facet 2 (literary content), on average, compared to those found using default Summon. Articles obtained using Social Sciences Abstracts also performed more favorably as measured by the average taxonomy score than those obtained using default Summon. Articles obtained using prescoped Summon also outperformed those obtained using default Summon on the average taxonomy score.

Figure 2 presents the average scores assigned using the relevance rubric. Articles chosen with default Summon received, on average, higher scores than those chosen with Social Sciences Abstracts, though this difference was only statistically significant at the less conservative $P < .05$ level.

TABLE 7
Average Relevance Rubric Scores by Article and Rater (score range 0–3)

	Default Summon				Prescoped Summon				Social Sciences Abstracts			
	Article 1		Article 2		Article 1		Article 2		Article 1		Article 2	
	Rater 1	Rater 2	Rater 1	Rater 2	Rater 1	Rater 2	Rater 1	Rater 2	Rater 1	Rater 2	Rater 1	Rater 2
Relevance Rubric	2.90	2.81	2.84	2.82	2.92	2.80	2.74	2.74	2.44	2.60	2.66	2.70

FIGURE 1
Taxonomy Scores for Each Tool

FIGURE 2
Relevance Score for Each Tool
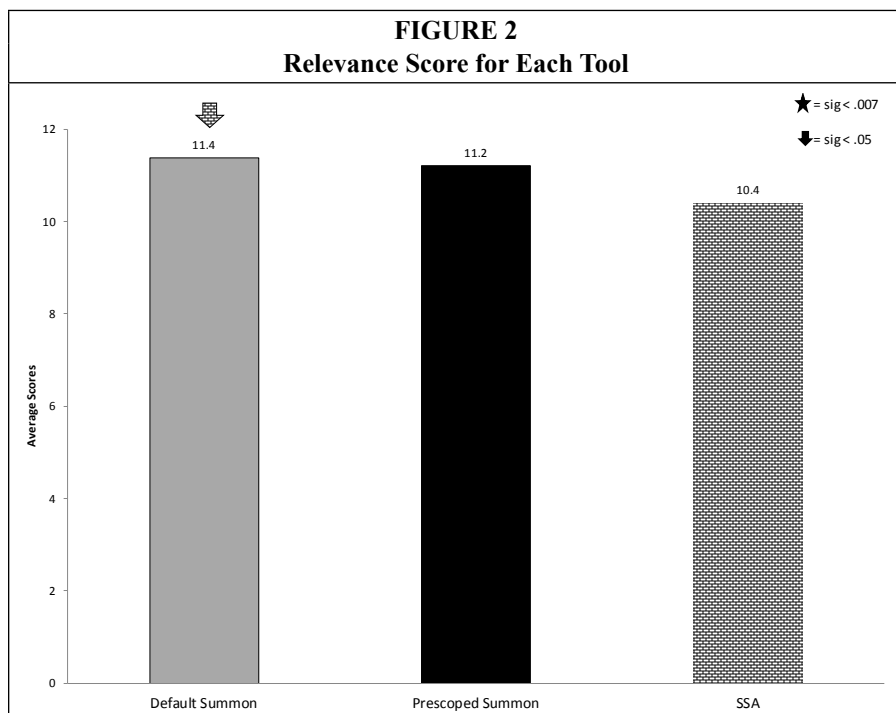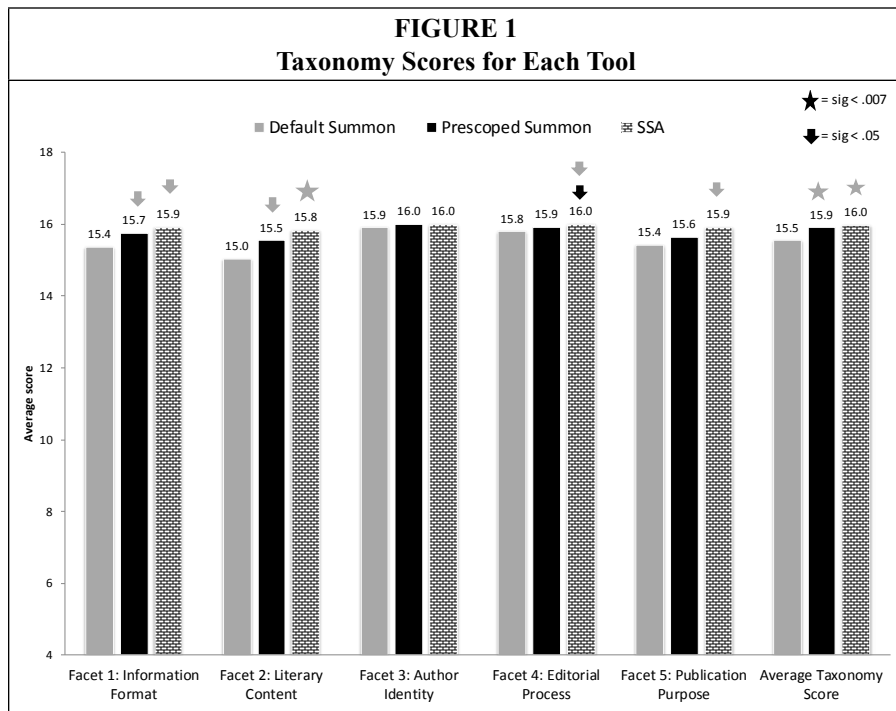
Using publication date as the indicator of article currency, we see in figure 3 that there are some differences between the earliest articles found in each tool as well as the average article dates. None of these differences are statistically significant, despite Social Sciences Abstracts' ranking of results defaulting to newest articles first and our instances of Summon prioritizing relevance in their ranking.

FIGURE 3
Minimum, Maximum, and Mean Publication Year by Tool
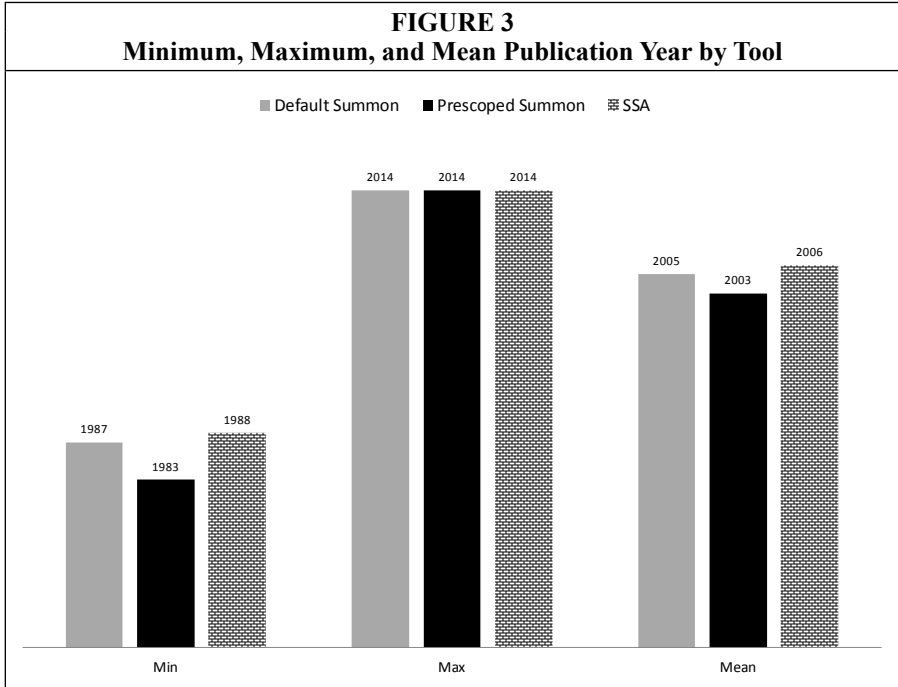
Effects of GPA and Library Instruction

We collected data from our student participants on their GPA and their previous exposure to library instruction with the intention of exploring whether either of these variables had an effect on their preference of search tool or the quality of articles selected. No statistically significant differences were found between participants who had received library instruction and those who had not. This suggests that library instruction has not had a demonstrable effect on either student preference for search tools or on the quality of articles that students select.

Similarly, no statistically significant differences were found to indicate that participants' preference for search tools was affected by their GPA. However, in measuring the correlation coefficients between GPA and the measures of article quality, a moderate positive correlation (0.319, $P = 0.024$) was found between GPA and article relevance in default Summon. This indicates that participants with higher GPAs were somewhat more likely to select more relevant articles from this search tool.

Discussion
Student Preferences for Search Tools

Our findings reinforce what other studies[40] have found with regard to student preference of search tools: that students tend to prefer discovery layers over traditional databases. Although the search interfaces for default Summon and prescoped Sum-

mon were the same single search box leading to a similar initial results page, students reported greater ease of use for default Summon (54% with highest ranking) than for prescoped Summon (38% with highest ranking). Further investigation would be needed to determine why this is the case; however, based on participant comments, we suspect that some students were turned off by the preselection of filters and may have assigned lower ratings for that reason.

When asked about their satisfaction with the search results, both of the Summon iterations got higher ratings from students than Social Sciences Abstracts. Interestingly, this corresponds with the lower relevance scores given by librarian raters to articles selected in Social Sciences Abstracts. When participants were asked why they chose each article, their answers inevitably focused on the article's relevance to their topic. It should not be surprising, then, that if students were finding less relevant articles in Social Sciences Abstracts (as indicated by the relevance scores), they would be less satisfied with the search results from that tool.

Student responses to the more general question of whether they would find each tool useful in their coursework also indicated a greater preference for default Summon than for prescoped Summon or Social Sciences Abstracts. When asked a related question, which tool they would prefer to use the most, answers were evenly split between Summon (n = 17), prescoped Summon (n = 16), and Social Sciences Abstracts (n = 17). This means that two-thirds of students had the greatest preference for a discovery layer.

Quality of Articles Selected by Student Participants
While our results suggest that most students prefer discovery layers for finding articles and have a stronger preference for the default settings rather than a prescoped version, the results are mixed with regard to which tool best supports students' ability to select high-quality articles.

Authority
It should be noted that the authority scores for articles selected from all three search tools were high. This is a testament to our participants' ability to select authoritative articles regardless of the search tool used, which may be a product of the population studied (upper-division undergraduates in a major where scholarly information is emphasized). Because scores were skewed toward the high end of the scoring range, the possibilities of finding statistically significant differences were slim. The fact that statistically significant differences between tools were found despite the skew indicates that these differences, while not numerically large, were consistent.

The taxonomy scores indicate that both Social Sciences Abstracts and prescoped Summon outperform default Summon with regard to article authority. As noted above, our library's default configuration of Summon includes nonscholarly information, but so do prescoped Summon (only newspapers were eliminated) and Social Sciences Abstracts, though in different proportions. We might have expected the upper-division students who constituted our sample to have a solid understanding of the importance of authority as a criterion for selecting "best quality" articles. In fact, one of the questions we asked students at the end of their searching sessions was to what extent they considered the credibility of the authors or the publication when selecting their articles. The vast majority of participants indicated that they had been looking for scholarly articles and assumed that these would be sufficiently authoritative.

We may conclude, then, that even when students have the intention of finding scholarly information, the authority of the sources they select is affected by the search tool used. One might imagine that the choice of search tool would have an even greater effect on students who had less exposure to the importance of scholarly information for

their coursework or who were unclear on what constitutes a scholarly article (perhaps lower-division students), and this might be a direction for future research.

It is also interesting to note the effect that the configuration of the discovery layer has on articles selected by students. As noted earlier, libraries have options in determining how their discovery layers are configured, including which information formats are included in the initial set of search results. Academic libraries may have good reason to include nonscholarly formats such as newspapers and magazines in the default results, as these types of articles may give students valuable background information on a topic before delving into the scholarly literature. These considerations will have to be weighed against the potential benefits of limiting the default results to scholarly formats, namely that students are more likely to select authoritative articles using such a configuration.

Relevance

We have described our inability to locate a published relevance rubric to meet the needs of this study and the difficulty determining relevance without the context provided by the narrative of a research paper or essay. The interrater reliability for our locally developed relevance rubric was lower than that for the taxonomy, though it was still relatively good. For these reasons, we are more cautious about our findings with regard to article relevance.

If we assume that our results are an accurate reflection of reality rather than a result of measurement error, we can say that default Summon outperformed Social Sciences Abstracts on article relevance, but only with the more liberal standard of statistical significance ($P < 0.05$). Social Sciences Abstracts indexes far fewer journals than Summon, so it may not be surprising that more relevant articles were present in Summon's results. That students would be able to find these more relevant articles among Summon's very long list of search results is either a testament to Summon's relevance ranking or to the ability of participants to parse result lists.

Interestingly, a slight correlation was found between students with higher GPAs and higher relevance scores for the articles selected in Summon. It may be that students with greater academic preparedness are better able to select relevant articles from a larger set of results.

Currency

No statistically significant differences between search tools were found with regard to the publication dates of the articles selected. Each tool's default ranking of search results no doubt affects student choice, given that students are unlikely to change default search settings[41] and often engage in satisficing behaviors.[42] The default ranking for Social Sciences Abstracts is newest items first, in contrast to Summon, which we have configured to prioritize results by relevance. Had these been configured differently, our results would likely have varied. We have no reason to conclude, therefore, that any of these tools has an effect on the currency of articles selected by students that is independent of the default configuration of search results ranking.

Participants were asked at the end of their searching to what extent they had considered the publication date when they chose their articles, and most students indicated an awareness that this was a consideration. Many mentioned that, for their course-related research, they tend to choose articles published in the last 10 years unless their topic dictates a broader timeframe.

Limitations of Study

As previously noted, the sample of participants was purposefully drawn from a particular major that we identified as a potential audience for a database such as Social

Sciences Abstracts. Additionally, we selected a sample of convenience from students in the Social and Behavioral Sciences major, meaning that our findings may not be entirely representative of this group and may not be generalizable to other contexts.

In this study, we were interested in exploring the differences in two broad categories of search tools: discovery layers and subject-specific indexing and abstracting databases. Practical considerations dictated that we choose only one tool to represent each category (plus a second configuration of the discovery layer). The tools that we chose may or may not be fully representative of these broader categories.

The statistical analysis of our data was limited by two factors: a relatively small sample size (n = 50) and the skewed distribution of the taxonomy scores. Both of these factors made it more challenging to find statistically significant results; but, despite these challenges, significant results were found. If the sample size had been larger or the results less skewed, additional statistically significant results may have been identified.

Conclusion

Rather than a strictly academic pursuit, we initiated this investigation anticipating that the results would provide actionable data for our library and perhaps for other academic libraries as well. The results of our study confirm an assumption that we, and likely many other librarians, held: that subject-specific indexing and abstracting databases still play an important role for libraries that have adopted discovery layers. Discovery layers and subject-specific indexing and abstracting databases have different strengths and can complement each other within the suite of library resources. Discovery layers may be a less intimidating way to introduce new students to searching for information, and they can be useful for finding known citations and conducting searches on esoteric topics. Subject-specific indexing and abstracting databases provide a less overwhelming set of search results as well as better options for advanced searching within a discipline, making them of particular use to upper-division undergraduates, graduate students, and faculty.

While the ideal combination of library search tools will depend heavily on local needs, there are a few lessons we have taken away from our research that may be of use for other libraries faced with decisions about which search tools to retain in times of limited financial resources.

Not all students prefer discovery tools. While the majority of our student participants preferred to search for articles using one of the Summon iterations, one-third of them indicated a preference for Social Sciences Abstracts, and this preference did not correlate with whether students had received prior library instruction that could have introduced them to this database. Providing a variety of search tools for students with different learning styles and different research needs allows students to choose a tool that best meets their needs. Instruction on which search tool may perform best under different circumstances would enable students to make informed decisions.

The tools that students prefer may not be those that give them the best results. Social Sciences Abstracts and prescoped Summon led students to choose more authoritative articles than our library's default configuration of Summon, though the latter came out ahead in student preference. This may be a compelling reason for libraries to not only maintain subscriptions to subject-specific indexing and abstracting databases, but to recommend them to students in reference and instruction interactions, especially upper-division and graduate students. This has implications for the importance of academic librarians in fostering student success: without proper guidance, students may be more likely to use what they perceive as the easiest tool regardless of its appropriateness for the task at hand.

Default configuration matters. The choices libraries make about their discovery layer's settings and filters affect the information sources that students will select. Libraries must make tough decisions about what configuration will work best for the majority of their users, weighing factors such as authority and scope. Research guides, which are generally subject-specific, may be a good location for discovery layer instances that are prescoped by subject and format. These could complement a broader discovery layer configuration on the library's homepage.

As of this writing, it is premature to predict all of the actions our library will take in response to these findings. It is safe to say that the results of this study will inform future discussions about which databases and search tools to retain in times of budget constraints. After a brief hiatus, we have renewed our subscription to Social Sciences Abstracts, and we have begun to consider whether our default configuration of Summon is best meeting the needs of our students. We have also discussed the possible utility of including on our online subject and course guides a Summon search widget that is prescoped to meet the subject and format needs of various disciplines.

Further Research

Many questions remain with regard to measuring relevance and the roles that different search tools play in leading students to relevant resources. While our results suggest that the broad indexing of discovery layers may give them an advantage in leading students to more relevant articles, this deserves further investigation. Similar investigations could be pursued to determine whether our findings hold true for different search tools or different user groups. The effects of library instruction and academic preparedness on search tool preference and selection of quality articles would also benefit from further research. The effects of library instruction, in particular, could be a fruitful topic of investigation. While we did not find instruction to have a significant effect on student satisfaction with search tools or their ability to find higher quality sources, further research in this area could inform instructional practice and speak to the value of academic librarians.

Acknowledgements

APPENDIX A. Search Prompts

Each participant was assigned one of the following tasks:
1. You are writing a research paper on the effects of children's toys on gender stereotypes. Find two of the best quality articles to use.

OR

2. You are writing a research paper on the factors that affect the academic achievement of children of immigrants. Find two of the best quality articles on this topic.

APPENDIX B. Post-Searching Survey for Student Participants

Participant code _____

Year in school _____

Are you a transfer student from another college or university? (yes/no)

Major and concentration _____GPA _____

How many library instruction sessions have you attended at CSUMB? _____

Have you had a one-on-one capstone consultation with a librarian? _____

How easy to use was each search tool?
Please think about each search tool in the order that you used it (first, second, third).
   First search tool: Learning to use this tool was easy.
   Do you agree with this statement?
   (strongly disagree) 1 2 3 4 5 (strongly agree)

   Second search tool: Learning to use this tool was easy.
   Do you agree with this statement?
   (strongly disagree) 1 2 3 4 5 (strongly agree)

   Third search tool: Learning to use this tool was easy.
   Do you agree with this statement?
   (strongly disagree) 1 2 3 4 5 (strongly agree)

How satisfied were you with the search results in each tool?
Please think about each search tool in the order that you used it (first, second, third).
   First search tool: How satisfied were you with the search results?
   (not satisfied) 1 2 3 4 5 (very satisfied)

   Second search tool: How satisfied were you with the search results?
   (not satisfied) 1 2 3 4 5 (very satisfied)

   Third search tool: How satisfied were you with the search results?
   (not satisfied) 1 2 3 4 5 (very satisfied)

Would you find these tools useful in your coursework?
Please think about each search tool in the order that you used it (first, second, third).
   First search tool: I would find this tool useful in my coursework.
   Do you agree with this statement?
   (strongly disagree) 1 2 3 4 5 (strongly agree)

   Second search tool: I would find this tool useful in my coursework.
   Do you agree with this statement?
   (strongly disagree) 1 2 3 4 5 (strongly agree)

Third search tool: I would find this tool useful in my coursework.
Do you agree with this statement?
(strongly disagree) 1 2 3 4 5 (strongly agree)

Which search tool did you prefer?
Please think about each search tool in the order that you used it (first, second, third)
Which tool would you most prefer to use?
First search tool / Second search tool / Third search tool

Which tool would you least prefer to use?
First search tool / Second search tool / Third search tool

**Please share any comments you have about the search tools or your
searching experience.**

APPENDIX C. Relevance Rubric

Score	3	2	1
Relevance	Source directly addresses one or more aspects of the topic	Source partially addresses an aspect of the topic	Source addresses topic in a very limited way

A score of 0 for articles not addressing the topic was also permitted, as were half points for articles falling in between categories.

APPENDIX D. Procedures for Condensing Variables

As noted above, rather than compare all 28 variables measured for each of the search tools indicated in tables 6 and 7, we pursued procedures to condense the number of variables. The first step involved collapsing the ratings assigned within raters. That is to say, for each rater, we summed the scores assigned each of the seven measures across articles 1 and 2. We repeated this step three times, once for pairs of articles generated using each search tool. Before doing so, we ran statistical tests to determine if there was any difference in the way articles 1 and 2 were rated within each rater. More specifically, we ran 14 separate paired sample t-tests. The first paired t-test addressed whether the average rating on facet 1 observed on article 1 rated by rater 1 ($M = 3.82$) was different than the average rating on the facet 1 for article 2 also rated by rater 1 ($M = 3.86$). We repeated this procedure 13 more times for each rater.

To control for inflated chance of obtaining statistically significant results due to the large number of comparisons being run, we used an adjusted P value of .007 (.05/7 comparisons) to establish statistical significance. The results of these analyses suggest that there was no difference in the scores that raters assigned articles 1 and 2 for each of the seven measures. As a result, we summed the ratings assigned each of the seven measures across both articles for each rater separately, yielding 14 variables per search tool instead of 28, as represented in table 8.

TABLE 8
Average Scores for Each Facet by Rater and Tool

	Default Summon		Pre-Scoped Summon		SSA	
	Rater 1	Rater 2	Rater 1	Rater 2	Rater 1	Rater 2
Facet 1	7.60	7.76	7.86	7.88	7.96	7.96
Facet 2	7.52	7.48	7.80	7.74	7.94	7.94
Facet 3	7.97	7.93	8.00	8.00	8.00	8.00
Facet 4	7.92	7.86	7.96	7.96	8.00	8.00
Facet 5	7.67	7.73	7.72	7.92	7.96	7.96
Average Taxonomy Score	7.75	7.77	7.96	7.94	7.98	7.98
Relevance Rubric	5.73	5.63	5.66	5.54	5.10	5.30

Again, rather than compare the 14 variables in table 8 for default Summon to the 14 variables for prescoped Summon (resulting in a total of 42 comparisons), we conducted analyses to determine whether the values assigned by raters 1 and 2 could be combined. This would be true to the extent that the information captured by both raters was essentially the same. We ran seven paired sample t-tests to determine whether scores assigned by rater 1 on each of the seven measures were different from the scores assigned by rater 2. We repeated these seven tests for each of the three search tools.

For the reason described above, we used a P value of .007 as the criterion to judge statistical significance for these tests. The results of these tests suggest that there was no difference in the scores assigned by raters 1 and 2 for each of the seven measures with one exception. Among the articles found using prescoped Summon, rater 2 scored facet 5 more highly ($M = 7.92$, n = 50) than rater 1 ($M = 7.72$; n = 50). Although statistically significant, the difference between 7.92 and 7.72 was deemed small enough to combine into one score.

Combining the ratings across raters 1 and 2 within each search tool resulted in a total of seven variables per search tool, a more manageable number.

Notes

    1. Andrew D. Asher, Lynda M. Duke, and Suzanne Wilson, "Paths of Discovery: Comparing the Search Effectiveness of EBSCO Discovery Service, Summon, Google Scholar, and Conventional Library Resources," *College & Research Libraries* 74, no. 5 (2013): 464.
    2. Marshall Breeding, "The Future of Library Resource Discovery: A White Paper Commissioned by the NISO Discovery to Delivery (D2D) Topic Committee," *Council of Undergraduate Research* (Baltimore, Md.: NISO 2015), available online at www.niso.org/apps/group_public/download.php/14487/future_library_resource_discovery.pdf [accessed 25 November 2015].
    3. Asher, Duke, and Wilson, "Paths of Discovery," 464.
    4. Melissa A. Hofmann and Sharon Q. Yang, "'Discovering' What's Changed: A Revisit of the OPACs of 260 Academic Libraries," *Library Hi Tech* 30, no. 2 (2012): 257.
    5. Julia Gross and Lutie Sheridan, "Web Scale Discovery: The User Experience," *New Library World* 112, no. 5/6 (2011): 242.
    6. Lisa Rose-Wiles and Melissa M. Hofmann, "Still Desperately Seeking Citations: Undergraduate Research in the Age of Web-Scale Discovery," *Journal of Library Administration* 53, no. 2/3 (2013): 149.
    7. Asher, Duke, and Wilson, "Paths of Discovery," 474.
    8. Helen Georgas, "Google vs. the Library: Student Preferences and Perceptions When Doing Research Using Google and a Federated Search Tool," *portal: Libraries and the Academy* 13, no. 2 (2013): 165.
    9. Rose-Wiles and Hofmann, "Still Desperately Seeking Citations," 149.

10. Matthew B. Hoy, "An Introduction to Web Scale Discovery Systems." *Medical Reference Services Quarterly* 31, no. 3 (2012): 324.

11. Rose-Wiles and Hofmann, "Still Desperately Seeking Citations," 150.

12. Courtney Lundrigan, Kevin Manuel, and May Yan, "'Pretty Rad': Explorations in User Satisfaction with a Discovery Layer at Ryerson University," *College & Research Libraries* 76, no. 1 (2015): 48.

13. Asher, Duke, and Wilson, "Paths of Discovery," 476.

14. Gross and Sheridan, "Web Scale Discovery," 244.

15. Ibid.

16. Rose-Wiles and Hofmann, "Still Desperately Seeking Citations," 152.

17. David Howard and Constance Wiebrands, "Culture Shock: Librarians' Response to Web Scale Search," *Conference Proceedings, ALIA Information Online Conference* (Feb. 2011), 4, available online at http://ro.ecu.edu.au/ecuworks/6206/ [accessed 26 August 2014].

18. Gross and Sheridan, "Web Scale Discovery," 242.

19. Rose-Wiles and Hofmann, "Still Desperately Seeking Citations," 152; Asher, Duke, and Wilson, "Paths of Discovery," 464.

20. Asher, Duke, and Wilson, "Paths of Discovery," 464.

21. Bill Kules and Robert Capra, "Influence of Training and Stage of Search on Gaze Behavior in a Library Catalog Faceted Search Interface," *Journal of the American Society for Information Science and Technology* 63, no. 1 (2012): 117.

22. Donald Owen Case, *Looking for Information: A Survey of Research on Information Seeking, Needs and Behavior* (Bingley, U.K.: Emerald Group Publishing, 2012).

23. Alison J. Head, "Project Information Literacy: What Can Be Learned about the Information-Seeking Behavior of Today's College Students?" *Association of College and Research Libraries (ACRL) Proceedings* (Chicago: ALA, 2013), 476.

24. Denise E. Agosto, "Bounded Rationality and Satisficing in Young People's Web-Based Decision Making," *Journal of the American Society for Information Science and Technology* 53, no. 1 (2002): 23.

25. Asher, Duke, and Wilson, "Paths of Discovery," 465.

26. Sean Cordes, "Student Perceptions of Search Tool Usability," *Internet Reference Services Quarterly* 19, no. 1 (2014): 4.

27. Gross and Sheridan, "Web Scale Discovery," 238.

28. Cordes, "Student Perceptions," 20.

29. Gross and Sheridan, "Web Scale Discovery," 242.

30. Asher, Duke, and Wilson, "Paths of Discovery," 466.

31. Karen Hovde, "Check the Citation: Library Instruction and Student Paper Bibliographies," *Research Strategies* 17, no. 1 (2000): 3.

32. Chris Leeder, Karen Marker, and Elizabeth Yakel, "A Faceted Taxonomy for Rating Student Bibliographies in an Online Information Literacy Game," *College & Research Libraries* 73, no. 2 (2012): 116.

33. Ibid.

34. Asher, Duke, and Wilson, "Paths of Discovery," 474; Kules and Capra, "Influence of Training," 115.

35. Leeder, Marker, and Yakel, "A Faceted Taxonomy," 117.

36. Ibid.

37. Lundrigan, Manuel, and Yan, "Pretty Rad," 47–49; Asher, Duke, and Wilson, "Paths of Discovery," 476.

38. Leeder, Marker, and Yakel, "A Faceted Taxonomy," 129; Asher, Duke, and Wilson, "Paths of Discovery."

39. Ibid.

40. Lundrigan, Manuel, and Yan, "Pretty Rad," 48; Asher, Duke, and Wilson, "Paths of Discovery," 476.

41. Asher, Duke, and Wilson, "Paths of Discovery," 464.

42. Agosto, "Bounded Rationality and Satisficing," 23.